은닉 마르코프 모델에 기반한 정상행위의 순서적 이벤트 모델링을 통한 침입탐지 시스템

최종호, 조성배
연세대학교 컴퓨터과학과

An Intrusion Detection System with Temporal Event Modeling based on Hidden Markov Model

Jongho Choy and Sung-Bae Cho
Computer Science Department, Yonsei University

요 약

사회분야 전반이 전산화되면서 전산시스템에 대한 효과적인 침입방지와 탐지가 중요한 문제로 대두되었다. 침입행위도 정상사용행위와 마찬가지로 전산시스템 서비스를 사용하므로 호출된 서비스의 순서로 나타난다. 본 논문에서는 정상사용행위에 대한 서비스 호출순서를 모델링한 후 사용자의 사용패턴을 정상행위와 비교해서 비정상행위(anomaly)를 탐지하는 접근방식을 사용한다. 정상행위 모델링에는 순서정보를 통계적으로 모델링하고 평가하는데 널리 쓰이고 있는 HMM(Hidden Markov Model)을 사용하였다. Sun사의 BSM 모듈로 얻어진 3명 사용자의 사용로그에 대하여 본 시스템을 적용한 결과, 학습되지 않은 u2r 침입에 대해 2.95%의 false-positive 오류에서 100%의 탐지율을 보여주었다.

1. 서 론

최근 급속한 정보통신 기반구조의 확산에 힘입어 컴퓨터를 이용한 정보처리에 새로운 장이 열리고 있으며, 그와 더불어 정보보안에 대한 문제가 심각하게 대두되고 있다. 실제로 미국 퍼듀대학의 보고에 의하면 지난 5년동안 정보에 대한 공격 위협이 250% 증가하여 그 피해액이 1천억 불에 달한다고 하며[7], 국내의 경우에도 정보보호센터의 조사에 의하면 네트워크를 통한 침입의 회수가 90년대 들어 급격히 늘어나는 추세이다[9]. 특히 최근 금융망이나 국방망, 전력망 등에 침입하는 사례가 늘고 있어, 불법적인 침입을 사전에 탐지하여 국가의 중요 정보통신 기반구조에 가해지는 피해를 차단할 필요가 있다.

침입탐지 시스템은 불법적인 사용이나 오용, 남용 등에 의한 침입을 알아내는 것으로[2, 4], 단일 컴퓨터는 물론이고 네트워크로 연결된 여러 컴퓨터를 감독할 수 있다. 이러한 시스템은 기본적으로 감사기록, 시스템 테이블, 네트워크 부하기록 등의 자료로부터 사용자의 행위에 대한 정보를 분석하는 작업을 한다. 이제까지 연구·개발된 침입탐지 방법은 공격행위의 정보를 이용하는지 정상행위의 정보를 이용하는지에 따라서 오용탐지와 비정상행위 탐지로 나눌 수 있다.

시스템의 알려진 공격행위들을 가지고 있다가 오용에 대한 침입을 탐지하는 오용 침입탐지 방법은 침입이 아닌데 침입이라 판정하는 오류(false-positive error)가 매우 적고 상대적으로 구현비용이 저렴하다는 장점이 있지만, 공격에 대한 정보를 계속 수집하는데 어려움이 있고 알려진 공격기법에 대해서만 탐지할 수 있다는 한계가 있다. 반면에 정상적인 시스템 사용에 대한 프로파일을 모델링하고 이에 벗어나는 행위를 탐지하는 비정상행위 탐지 방법은 정상행위에 대한 대량의 데이터를 분석하여야 하기 때문에 구현비용이 크기는 하지만, 알려지지 않은 새로운 공격도 탐지할 수 있어 침입을 침입이 아니라고 판정하는 오류(false-negative error)를 줄일 가능성이 있다.

침입탐지 시스템은 사용자의 키입력, 시스템호출, 접속당 사용시간, 시스템의 평균부하 등 다양한 관찰심볼을 기반으로 침입여부를 결정한다. 침입탐지문제는 관찰된 사용패턴을 침입행위와 정상행위로 분류하는 문제로 생각할 수 있다. 컴퓨터에서 사용자의 행위시퀀스는 일반적으로 몇 가지 정해진 패턴을 따르며 명령어 수행이나 시스템 호출 등의 이벤트로 관찰된다. 따라서 사용자의 행위시퀀스에 기반해서 사용자의 행위를 설명하고 특성을 파악할 수 있는 모델을 구축할 수 있다면 이를 기반으로 정상행위를 모델링한 후 사용자의 행위가 정상행위인지 판정할 수 있다. 본 논문에서는 음성인식 및 여러 분야에서 알려지지 않은 대상을 모델링하는데 널리 쓰이고 있는 HMM을 사용자의 정상행위 모델링과 비정상행위의 판정에 적용하고자 한다.

2. 관련연구

오용탐지기법이 알려진 오용행위만을 탐지할 수 있다는 한계가 있기 때문에 날로 다양화되어 가는 침입기법에 대응하기 위해 비정상행위 탐지기법에 관한 연구가 활발해지고 있다. 이들은 사용자의 사용패턴에 기반하거나[3] 프로그램의 사용패턴에 기반하여[1, 8] 정상행위를 모델링한 후 판정하려는 순서적 이벤트를 여러 기법을 통하여 모델링된 정상행위와 비교해서 정상행위에서 얼마나 벗어나는 행위인지를 측정한다.

[3]은 사용자별로 사용패턴 벡터를 구축한 후 판정하려는 이벤트

시퀀스 벡터와의 유사성을 패턴매칭을 통해 판별하며, [1, 8]은 프로그램별 사용패턴을 모델링한다. [1]은 패턴매칭, BP 신경망, Elman 신경망을 사용한 경우를 비교하고 있으며, [8]은 빈도수, 데이터 마이닝기법, HMM의 성능을 비교한다. [1, 8]에서 연구결과 이벤트들간의 순서적 정보를 이용한 경우에 다른 경우보다 나은 성능을 보여주었으며 [3]의 경우에도 패턴매칭시 순서적 정보를 중요한 요인으로 고려하고 있다.

3. HMM을 사용한 침입탐지

3.1 침입탐지 시스템 개요

본 논문에서 개발하는 침입탐지시스템은 그림 1과 같이 데이터 필터링과 데이터 축약을 담당하는 전처리 모듈과 정상행위 모델링과 추론 및 판정을 담당하는 비정상행위 판정모듈로 구성된다. 모델학습을 통해 정상행위를 프로파일 데이터베이스로 구축하여 판정시 사용한다.

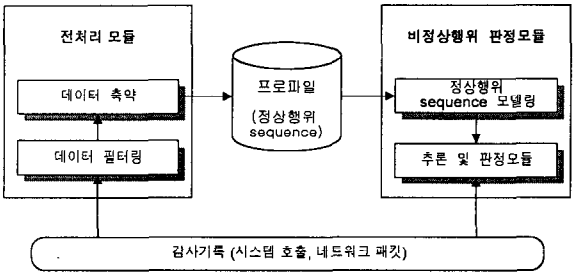

그림 1 . 침입탐지 시스템 개요

3.2 전처리

사용자 감사기록으로는 사용자의 시스템 호출번호를 사용하였다. 모든 시스템호출번호가 다 사용되지 않으므로 통계적으로 빈도가 높은 49개의 시스템호출에 대해 0부터 48번까지의 번호를 부여하였고 그 밖의 시스템 호출은 49번을 부여하여 총 50개의 축약된 시스템 호출번호를 사용하였다. 순서적으로 생성되는 시스템 호출번호는 일정 크기의 윈도우를 옆으로 이동시켜가면서 윈도우 크기 만한 시퀀스로 추출하였다.

번호	호출	번호	호출	번호	호출
0	exit	17	mkdir	34	munmap
1	fork	18	rmdir	35	setegid
2	create	19	setrlimit	36	seteuid
3	link	20	pathconf	37	putmsg
4	unlink	21	open_r	38	getmsg
5	chdir	22	open_w	39	audition_setcond
6	chmod	23	open_wc	40	statvfs
7	chown	24	open_rw	41	sysinfo
8	kill	25	open_rwc	42	fork1
9	symlink	26	close	43	sockconnect
10	readlink	27	getaudit	44	login
11	execve	28	setaudit	45	logout
12	vfork	29	ioctl	46	telnet
13	getgroups	30	setuid	47	rlogin
14	setpgrp	31	utime	48	su
15	fcntl	32	nice	49	etc.
16	rename	33	setgid		

표 1 . 축약된 시스템호출

3.3 침입탐지를 위한 HMM

HMM은 실제적인 생성모델을 알 수 없고 단지 생성된 시퀀스에 의해서만 확률적으로 관측할 수 있는 이중으로 확률적인 절차로서[5, 6], 사용자의 행위시퀀스를 모델링하기에 유용한 도구이다. HMM은 관찰 시퀀스의 길이, 상태수, 심볼수와 학습에 의해 조정되는 전이확률, 관측확률, 초기상태분포로 구성이 된다. 전이확률은 한 상태에서 다음상태로 전이할 확률을 나타내며, 관측확률은 한 상태에서 특정 심볼이 관측될 확률을 나타낸다. 초기 상태 분포는 처음에 해당 상태에서 시작할 확률을 나타낸다. HMM은 다음과 같이 표현되며, 모델 λ는 간략히 (A, B, π)로 표현될 수 있다.

- T : 관찰 시퀀스의 길이
- N : 모델의 상태수
- M : 관찰 심볼의 수
- $Q = q_1, q_2, \ldots, q_N$: 상태들
- $V = v_1, v_2, \ldots, v_M$: 가능한 관찰심볼의 이산적인 집합
- $A = \{a_{ij}\}, a_{ij} = \Pr(q_j \text{ at } t+1 | q_i \text{ at } t)$: 상태전이 확률분포
- $B = \{b_j(k)\}, b_j(k) = \Pr(v_k \text{ at } t | q_j \text{ at } t)$: 관측 심볼 확률분포
- $\pi = \{\pi_i\}, \pi_i = \Pr(q_i \text{ at } t=1)$: 초기 상태 분포

가) 정상행위 모델링

정상행위 모델링은 전처리 단계에서 생성된 정상행위 시퀀스를 기반으로 HMM의 매개변수를 결정하는 과정이다. HMM의 매개변수 결정은 주어진 시퀀스 O가 해당 모델 λ로부터 나왔을 확률인 $\Pr(O|\lambda)$값이 최대가 되도록 $\lambda = (A, B, \pi)$를 조정한다. 이를 계산하는 해석적인 방법은 알려져있지 않고 반복적으로 λ를 결정하는 방법으로 Baum-Welch의 재추정식이 있다[5, 6].

나) 비정상행위 판정

비정상행위 판정에서는 이미 구축되어 있는 정상행위별 HMM에 사용자행위시퀀스를 입력으로 넣어 각 정상행위에서 현재 행위가 생성되었을 확률을 구한다. 확률을 구하는 방법으로는 forward-backward procedure나 Viterbi 알고리즘을 사용할 수 있다[5, 6]. 각 모델별로 구해진 확률은 판정모듈에 전달되어 비정상행위인지 판정한다.

4. 실험결과

실험 데이터로는 3명의 사용자가 1주일간 발생시킨 데이터를 사용하였다. 주사용 프로그램은 문서편집기와 컴파일러, 그리고 사용자가 작성한 프로그램이었다. 학습 데이터와 테스트 데이터 모두 10,000개를 사용했으며, 테스트 데이터에는 각 사용자별로 u2r 침입을 17차례 넣었다. 실험은 1차로 전체 사용자 데이터를 단일 모델을 사용해서 수행하였다. 그 후 각 사용자별로 별도의 모델을 가정해서 2차 실험을 수행하였다. HMM의 매개변수로는 상태수 10, 심볼수 50, 시퀀스 길이 30을 적용하였고 모델은 left-to-right 모델을 사용했다.

그림 3은 사용자별 모델링시 사용자3에 침입행위가 발생한 경우를 보여준다. 침입행위가 시작된 시간 11에서부터 침입행위가 종료된 시간 32사이에 있는 시퀀스들의 평가값이 급격히 낮아져 HMM이 효과

적으로 침입을 탐지하고 있는 것을 보여주고 있다.

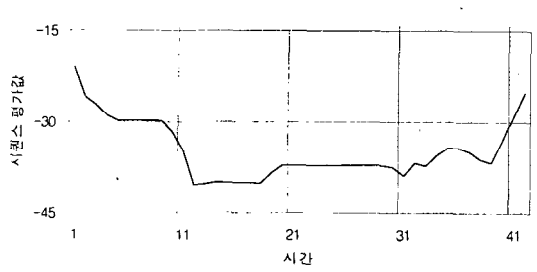

그림 2 . 침입행위 발생시 시간에 따른 평가값의 변화

표 1은 각 모델별 탐지율과 false-positive 오류율을 보여준다. 모든 테스트 데이터를 단일 모델로 모델링한 경우보다 각 사용자별로 모델링한 경우에 전반적으로 더 낮은 false-positive 오류율을 보여준다.

	단일 모델링	사용자별 모델링		
		사용자 1	사용자 2	사용자 3
정상행위 평균(편차)	-14.74 (10.42)	-15.54 (5.88)	-4.72 (8.42)	-14.25 (12.34)
임 계 값	-32.45	-26.21	-31.35	-36.92
탐 지 율	100%	100%	100%	100%
false-positive 오류율	2.95%	4.31%	1.73%	1.96%

표 2 . 비정상행위 탐지결과

그림 4는 ROC(Receiver Operating Characteristic) 곡선으로서 [1] 변경 가능한 매개변수의 조정에 따른 탐지율과 false-positive 오류율의 변화를 보여주고 있다. 본 곡선에서는 임계값을 조정해서 ROC 곡선을 얻었다. 바람직한 침입탐지시스템은 낮은 false-positive 오류에서 높은 침입탐지율을 보여주어야 하므로 곡선이 왼쪽에 있을수록 좋은 성능을 나타낸다. 이 경우에도 단일 모델링의 경우보다 세분화된 모델링의 경우가 전반적으로 좋은 성능을 보여주었다.

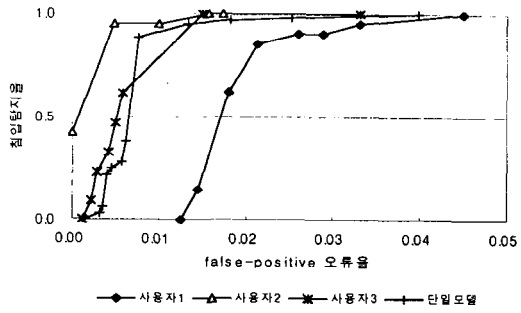

그림 3 . ROC 곡선

5. 결론 및 향후연구

본 논문에서는 HMM을 사용하여 사용자의 정상행위를 모델링한 후 순서적으로 생성되는 이벤트를 분석하여 비정상행위를 판정하는 컴퓨터 침입탐지기법을 제안하였다. 본 기법은 침입행위에 대해 2.95% 의 false-positive 오류에서 100 %의 탐지율을 보여주었으며, false-positive 오류가 발생한 부분은 학습시 정상행위로 모델링되지 않은 것이었다. 따라서, 포괄적으로 정상행위 데이터를 제공할 수 있다면 HMM이 효과적인 침입탐지기법으로 사용될 수 있음을 알 수 있었다.

또한 하나의 모델로 모델링한 것보다 사용자별로 모델링한 경우가 오류가 더 적음을 확인하였다. 따라서 보다 효과적인 모델의 수를 결정할 필요가 있다. 앞으로 사용자의 지위와 역할, 그리고 사용패턴에 기반한 다양하고 포괄적인 정상행위 추출 기법과 사용자의 행위를 효과적으로 나타낼 수 있는 변별력있는 사용자 이벤트 추출에 관한 연구가 수행되어야 할 것이다.

참고문헌

[1] A. K. Ghosh, A. Schwartzbard and M. Schatz, "Learning program behavior profiles for intrusion detection," *Proc. Workshop on Intrusion Detection and Network Monitoring*, Santa Clara, USA, April 1999.

[2] S. Kumar and E. H. Spafford, "An application of pattern matching in intrusion detection," *Technical Report CSD-TR-94-013*, 1994.

[3] T. Lane and C. E. Broadly, "Temporal sequence learning and data reduction for anomaly detection," *Proc. ACCS '98*, pp. 150~158, 1997.

[4] T. F. Lunt, "A survey of intrusion detection techniques," *Computer & Security*, vol. 12, no. 4, June 1993.

[5] L. R. Rabiner and B. H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, pp. 4~16, January 1986.

[6] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257~286, Febebuary 1989.

[7] E. H. Spafford, *Security Seminar*, Department of Computer Science, Purdue University, Jan 1996.

[8] C. Warrender, S. Forrest and B. Pearlmutter, "Detecting intrusions using system calls: Alternative data models," *Proc. IEEE Symposium on Security and Privacy*, May 1999.

[9] 한국정보보호센터, 호스트기반 침입탐지시스템 개발에 관한 연구, 1998년 12월.